Method of Identification of Proteins

Field of the Invention

The present invention relates to biological molecules or biomolecules. In particular, the present invention relates to the identification and/or quantification of biomolecules.

Background of the Invention

The development of reliable methodologies to rapidly identify and quantify large numbers of proteins is a central challenge in the emerging field of proteomics. At present, two-dimensional gel electrophoresis (2DE) combined with mass spectrometry (MS) protein identification is the workhorse platform for proteomic research, due to its ability to resolve and quantify complex protein mixtures (Aebersold, 2003).

However, intrinsic limitations of the 2DE/MS approach include the existence of significant gel to gel variability, and the possibility of several proteins co-migrating as one single spot, (Vuong et al, 2000; Campostrini et al, 2005) which can impede accurate quantitation and identification.

As an alternative to 2DE/MS, isotope-coded affinity tag (ICAT) labeling allows individual peptides in complex mixtures to be identified and quantitated (Gygi et al, 1999). Originally used in combination with liquid chromatography-tandem

mass spectrometry (LC-MS/MS), the ICAT reagent comprises a biotin affinity tag and thiol specific reactive group that labels peptides at cysteine residues (Vuong et al, 2000).

The mass spectra associated with ICAT, however, can be highly complicated as the ICAT targeted cysteine inhibits smooth analysis. Furthermore, peptide fragments can sometimes be formed due to the ICAT reagents during MS/MS analysis, resulting in protein misidentification (Haynes and Yates, 2000), and when coupled with on-line HPLC (high-performance liquid chromatography) separation, the same peptides with differential ICAT labels do not always co-elute, causing challenges for protein quantitation (Herbert et al, 2001). These findings thus raise the need for more robust approaches towards the reliable quantitation of protein differences in complex mixtures.

As such, improved means to identify and quantify large numbers of biomolecules such as proteins will be welcome.

Summary of the Invention

The present invention addresses the problems above and provides methods of identifying and/or quantifying proteins. In general, the present invention provides new method of identifying and/or quantifying large numbers of proteins. In particular, there is provided a method of identifying and/or quantifying biomarkers.

In one aspect, the present invention provides a method of identifying and/or quantifying at least one protein, the method comprising:

(a) providing at least one protein;

(b) labelling at least one amino acid residue of the protein(s) with at least one agent;

(c) separating the protein(s) based on a first characteristic;

(d) separating the protein(s) based on a second characteristic; and

(e) analysing the protein(s) by mass spectrometry, thereby identifying and/or quantifying the at least one protein.

There is also provided a method of identifying and/or quantifying at least one protein from at least two biological materials, the method comprising:

(a) providing a first biological material comprising at least one protein and at least one further biological material comprising at least one protein, wherein the protein of the first biological material and the protein of each further biological material are the same;

(b) labelling at least one amino acid residue of the protein(s) with at least one agent;

(c) separating the protein(s) based on a first characteristic;

(d) separating the protein(s) based on a second characteristic;

(e) analysing the protein(s) by mass spectrometry, thereby determining the expression of the at least one protein; and

(f) comparing the expression of the protein(s) in the at least two biological materials.

The at least one biological material may be in the form of a sample. The biological materials may be obtained from at least one subject suffering from, or diagnosed with, at least two different disease conditions. The at least two biological materials may be obtained from at least one subject under different treatment conditions. The at least two biological materials may be obtained from the same subject at different time points.

There is also provided a method of screening for at least one drug-responsive protein, the method comprising:

(a) administering a drug to at least one subject;

(b) obtaining at least one protein from the subject;

(c) labelling at least one amino acid residue of the protein(s) with at least one agent;

(d) separating the protein(s) based on a first characteristic;

(e) separating the protein(s) based on a second characteristic;

(f) analysing the protein(s) by mass spectrometry, and determining any variation in the expression of the protein(s), thereby identifying the at least one drug-responsive protein.

There is also provided a method of screening for at least one drug that causes a variation in expression of at least one protein, the method comprising:

(a) administering a drug to at least one subject;

(b) obtaining at least one protein from the subject;

(c) labelling at least one amino acid residue of the protein(s) with at least one agent;

(d) separating the protein(s) based on a first characteristic;

(e) separating the protein(s) based on a second characteristic;

(f) analysing the protein(s) by mass spectrometry, and determining any variation in the expression of the protein(s), thereby identifying the at least one drug.

The agent may be a sulfenyl halide. The agent may be 2-nitrobenzenesulfenyl chloride (NBS). The amino acid residue may be tryptophan. The separating steps may be based on at least one characteristic selected from the group consisting of mass, functional group and isoelectric point. The separating steps may be steps for two-dimensional gel electrophoresis (2DE), The analysing step by mass spectrometry may comprise ionizing the protein(s) and detecting the separated protein(s). The ionizing may be selected from the group consisting of electrospray ionisation (ESI), matrix assisted laser desorption/ionisation (MALDI), surface enhanced laser desorption/ionisation (SELDI), nanospray ionisation, atmospheric pressure chemical ionisation (APCI), chemical ionisation (CI), electron impact (EI), fast atom bombardment

(FAB), field desorption / field ionisation (FD/FI) and thermospray ionisation (TSP). The analysing may be by tandem mass spectrometry. The tandem mass spectrometry may use quadrupole-time-of-flight and/or ion trap geometries. The protein(s) may be biomarkers. The separating steps may be followed by digesting the protein(s) before the analyzing step. The digesting may be by at least one enzyme. The digesting may be performed on a membrane.

There is also provided a kit comprising at least one isotope of a sulfenyl halide compound. The sulfenyl halide compound may be 2-nitrobenzenesulfenyl chloride (NBS). The compound may be for labelling of biomolecules. The kit may be for identifying and/or quantifying at least one protein. The kit may further comprise packaging and information pertaining to the use of the compound for purposes such as diagnosis of disease states.

Brief Description of the Figures

Figure 1. (A) Modification of tryptophan residues with 2-nitrobenzensulfenyl chloride (NBS). * represents either ^{12}C (light) or ^{13}C (heavy). (B) Schematic representation of the NBS/2DE/MS protein identification/quantitation strategy. In this study, the selection of spots for excision was based on a prior analysis using conventional 2DE gels (see Results for details). However, it is also clearly feasible to excise all the observable spots on the NBS gel for subsequent analysis, if a more global proteomic survey is required.

Figure 2. (A) MALDI-TOF spectrum of tryptic peptides of chicken ovalbumin. Two pairs of NBS-labeled peptides (m/z 1734.7/1740.7 and m/z 2012.0/2018.0) are highlighted. (B) Mass spectra of the NBS-labeled peptides pair (m/z 1734.7/1740.7), showing the consistency between the amount of protein labeled and the resulting MALDI-TOF MS spectra.

Figures 3. (A) NBS-labeled peptide pair m/z 1135.5/1141.5 from protein Keratin-18 (K1C18_HUMAN) was acquired by MALDI-TOF MS using the CHCA/DHB matrices; (B) The same peptide pair was analysed using a combination of HNBA and CHCA/DHB matrices. To facilitate the comparison, the scale of the full range spectra (upper panels) was set to 100 mV, while the scale of the zoom-in spectra (lower panels) was set to 10 mV. The figures on the right show that HNBA in combination with CHCA/DHB greatly enhanced the ionisation of NBS-labeled peptides.

Figures 4. 2DE map of NBS-labeled protein spots that were excised and identified by MS analysis. Fifty-eight protein spots were positively identified by peptide mass fingerprinting from 88 spots with differential fluorescence stain intensity as determined by 2DE analysis. Among them, 44 protein spots were detected with NBS-labeled peptide pairs, which were quantified by calculating $^{13}C/^{12}C$ peptide volume ratios.

Figures 5. Correlation of relative protein quantitation (ratios of TAM and vehicle treated MCF-7 cells) by NBS/2DE/MS and 2DE approaches. Among the 44 differentially expressed proteins, the correlation coefficient between the ratios estimated by the two methods was 0.51, indicating a statistically significant correlation between these two protein quantitation methods. Spots 14 and 21 are highlighted for further concordance study (see Figures 6 and 7).

Figure 6. Spot 14 (disulfide isomerase ER-60) has good concordance between the NBS/2DE/MS and 2DE quantitation. PDQuest 3D image analysis on this spot reveals a clear single image peak. In contrast, Spot 21 (CK19) has poor concordance between the NBS/2DE/MS and 2DE quantitation. 3D image analysis indicates that there are at least 2 image peaks detected in the protein spot, indicating co-migration of proteins.

Figure 7. Mass spectra of tryptic peptides of Spot-21 (CK19). The peptide sequencing mass spectra of the matched (m/z 1222.75) and unmatched peptide (m/z 1419.85) are shown with their y-ions. The matched peptide 1222.75 (TKFETEQALR) (SEQ ID NO:54) was found from CK19, while the unmatched peptide m/z 1419.85 (LEGLTDEINFLR) (SEQ ID NO:55) was from CK8.

Figure 8. Two isoforms of the heat shock 60 kDa protein (CH60_HUMAN) with different isoelectric points (Spots 7 and 8). (A) Based on the NBS labeling

quantitation, this protein was up-regulated on Spot 8 but down-regulated on Spot 7. (B) The ratios of fluorescence labeling of 2DE spots were concordant with the NBS labeling ratios.

Figure 9. Technical Reproducibility of the NBS/2DE/MS Platform. The fold changes of protein quantity upon TAM treatment for duplicate samples from the same batch of MCF-7 cells were measured by NBS/2DE/MS. The result indicated a high reproducibility of the NBS labeling method. The correlation coefficient between the two experiments was 0.71, indicating a strong correlation. Twenty-one protein spots from two different batches of MCF-7 cells were quantified by NBS/2DE/MS. The correlation coefficient between the two independent batches was 0.57, showing a reasonably good reproducibility of protein quantitation between the batches.

Figure 10. Validation of NBS quantification by immunoblotting. (A) Two representative proteins GRP78 and CK19, which were found to change upon TAM treatment, were validated by immunoblotting with the protein specific antibodies. GRP78 is up-regulated in TAM treated MCF-7 cells, while CK19 protein quantity is considerably decreasing at 48hrs after TAM treatment. Blotting with anti-vinculin antibody represents equal loading of samples. (B) (Top) Immunoblotting reveals the down-regulation of PA2G4 protein at 24 and 48 hr after TAM treatment. Blotting with β-Actin antibody is shown for the equal loading of total cell lysates. (Below) RT-PCR amplification of PA2G4 and β-

actin transcripts from oligo-dT primed cDNA shows the relative down-regulation of PA2G4 mRNA upon TAM treatment.

Figure 11 (A, B, C). 2DE images of MCF-7 cells treated with TAM or vehicle for 48 hr, and the NBS-labeled combined TAM or vehicle treated cells. All the images show very similar protein patterns. The results of qualitative and quantitative image analyses are presented in Figure 12.

Figure 12. PDQuest image analysis results comparing the 2DE images of TAM treated or vehicle treated (untreated) MCF-7 cells. Triplicate gels were run for each treatment. (A) Match rates for the two sets of gels was ca. 85-95% among the ca. 340 visible gel spots; (B) Numbers of differentially displayed protein spots at qualitative and quantitative levels.

Detailed Description

Bibliographic references mentioned in the present specification are for convenience listed in the form of a list of references and added at the end of the examples. The whole content of such bibliographic references is herein incorporated by reference.

Definitions

Biological material – any material from a biological source. The material may comprise biomolecules, cells, tissue and/or fluid. The material may be obtained from cell culture, an animal, a mammalian or human subject.

Biomarker - Biomarkers are biological molecules (proteins, polypeptides, carbohydrates, lipids, etc, and their derivatives) associated with the presence and severity of specific disease states or different treatment conditions. Under the present invention, a protein or protein derivative may be used as a biomarker. Biomarkers are detectable and measurable by a variety of methods. To identify a biomarker, it is usually necessary to detect a difference or change in the expression, abundance or concentration of a biological molecule and identifying the particular molecule changed.

Expression or abundance of a biomarker – the expression of a biomarker may be determined from the presence or abundance of its gene, gene transcript, and gene product. The terms "gene" and "gene transcript" include RNA sequences complementary to the gene and cDNA sequences obtained by reverse transcription of the gene transcript. The terms also include the wild type gene, variations and mutations of the gene and gene transcript wherein the variation or mutation share substantial identity with the gene or gene transcript. Similarly, "gene product" includes wild type gene product, variations, fragments or derivatives thereof. "Substantial identity" means that the variants of the gene or gene product retain sufficient identity for them to be detectable by methods

and probes used for the wild type gene or gene product and retain the same functions as the wild type gene or gene products. The determination may be qualitative such as whether the biomarker is expressed, or the determination may be quantitative, or the determination may be semi-quantitative by any method known in the art such as by microarray technology or polymerase chain reaction. Departure (increase or decrease) from normal levels in the undiseased state can indicate a disease state or predisposition to a disease state. For example, the overexpression or increased abundance of a protein may be indicative of a disease state, the severity of the disease state and hence the prognosis for the subject in which the biomarker was determined, and/or it might be an indicator of susceptibility to a disease state. The difference in expression and abundance may be determined between different biomarkers or between the same biomarker under different conditions or time points. The expression of the biomarker(s) may then be compared and correlated to a reference value or to other values obtained at different time points or between different biomarkers to determine the correlation with the presence, severity of a disease state, or to determine the efficacy of a treatment, or to determine a prognosis of the disease outcome for the subject. Reference values may be determined from a statistically significant number of subjects suffering from or not suffering from. The expression of the biomarker(s) may also be used in conjunction with other suitable diagnostic or prognostic markers, biomarkers or indices to obtain a higher level of confidence. When a biomarker is said to be "over expressed" when compared

to controls, it is meant that the expression of that biomarker is at an abundance or level that is statistically significantly more than that naturally expressed by at least one wild type or non-mutant control subject not diagnosed with the disease state or condition. Similarly, when a biomarker is "under expressed", the expression of that biomarker is statistically significantly less than that naturally expressed by at least one wild type or non-mutant control subject not diagnosed with the disease state or condition. Under this definition, a subject genetically deficient for that biomarker cannot be said to be under expressing that biomarker. Similarly, when transfected with the gene for that biomarker so that that biomarker is expressed, the deficient subject cannot be said to be over expressing the biomarker as he was originally deficient in that biomarker.

Control – A reference subject, experiment or value by which values obtained in tests can be compared against. Control values or ranges usually represent the "normal" state so that a statistically difference or deviation of the control values or ranges represents an abnormal or disease state. A person skilled in the art will know how to select and/or obtain control subjects, experiments or values for use as references.

Selecting a subject or candidate – to determine the suitability of a subject or candidate for a clinical trial, experimentation, diagnostic and/or other tests by measuring at least one characteristic of the subject or candidate. The value representing the characteristic is then compared to a reference or control value or range of values, and the result is used to determine if the subject is suitable

for the trial, experimentation, diagnostic and/or other test. A person skilled in the art will know how to select subjects or candidates based on their amenability to a particular treatment, or their susceptibility to a particular challenge or disease.

Cell proliferation disorder – a condition in which there is abnormal growth of cells.

Cancer – A malignant and uncontrolled growth of cells in one part of the body that can spread to other parts of the body.

Nucleic acid — "Nucleic acids" are linear polymers of nucleotides, linked by 3', 5' phosphodiester linkages. In DNA, deoxyribonucleic acid, the sugar group is deoxyribose and the bases of the nucleotides adenine, guanine, thymine and cytosine. RNA or ribonucleic acid, has ribose as the sugar and uracil replaces thymine. An "isolated nucleic acid" is a nucleic acid the structure of which is not identical to that of any naturally occurring nucleic acid or to that of any fragment of a naturally occurring genomic nucleic acid. The term therefore covers, for example, (a) a DNA which has the sequence of part of a naturally occurring genomic DNA molecule but is not flanked by both of the coding sequences that flank that part of the molecule in the genome of the organism in which it naturally occurs; (b) a nucleic acid incorporated into a vector or into the genomic DNA of a prokaryote or eukaryote in a manner such that the resulting molecule is not identical to any naturally occurring vector or genomic DNA; (c) a separate molecule such as a cDNA, a genomic fragment, a fragment produced by polymerase chain reaction (PCR), or a restriction fragment; and (d) a

recombinant nucleotide sequence that is part of a hybrid gene, i.e., a gene encoding a fusion protein. The invention also features an isolated nucleic acid at least 50% (including any percentage between 50% and 100%, e.g., 85%, 95%, or 100%) identical to SEQ ID NO: 1, which can be used, e.g., for detection of the human PA2G4 gene in a subject, or screening of therapeutic compounds for treating a cell proliferation-associated disorder.

Gene product – Under the present invention, the gene product refers to a gene product other than a protein. The gene product may be a length of RNA with a biological functional other than a gene transcript coded for by a particular gene.

Gene transcript – the messenger RNA sequence coded for by a gene.

Protein - A biological molecule composed of one or more chains of amino acids in a specific order. Proteins may have derivatives such as isoforms. A protein isoform is a version of a protein with some small differences, usually a splice variant or the product of some post-translational modification. Under the present invention, a protein also encompasses fragments that are sufficiently large enough for the protein to be detected, identified and/or quantified by the method(s) used. A protein complex is a mixture of different proteins. The expression of proteins may be affected or influenced by a drug. Such proteins are called drug-responsive proteins.

Proteome – The set of proteins expressed by a cell or organ at a particular time and under specific conditions. Accordingly, proteomic analysis is the study of the full set of proteins encoded by a genome.

Antibody - An immunoglobulin protein produced by B-lymphocytes of the immune system that binds to a specific antigen molecule. The term includes monoclonal antibodies, polyclonal antibodies as well as fragments thereof, such as Fab, F(ab')2, and Fv fragments.

Drug – any compound used for the treatment (amelioration of, reduction of, or cure for) any disease or pathological state or condition.

Screening – to select a compound that has an effect in any disease or pathological condition. Screening may involve detecting, identifying and/or quantifying any biomolecule such as a protein, the expression of which is affected by the compound.

Treating - the administration of a composition or drug to a subject with a cell proliferation-associated disorder, with the purpose to cure, alleviate, relieve, remedy, prevent, or ameliorate the disorder, the symptom of the disorder, the disease state secondary to the disorder, or the predisposition toward the disorder. An "effective amount" is an amount of the composition that is capable of producing a medically desirable result, e.g., as described above, in a treated subject. This method can be performed alone or in conjunction with other drugs or therapy. Treating includes reduction of the biomarker to a level found in the non-disease state.

Hybridizing, hybridisable – a biomolecule such as a sequence of nucleic acids may hybridize to another sequence of nucleic acids if there are sufficient stretches of complementary nucleotides between them. For peptides and proteins, one protein such as an antibody or fragment thereof may hybridize to

another protein if there are sufficient complementary or recognition structures between these two biomolecules.

Blood derivative – Blood is made up of a cellular and a liquid component. A blood derivative is any of one or more of the blood components less than whole blood.

In general, the present invention provides new method of identifying and/or quantifying large numbers of proteins. In particular, there is provided a method of identifying and/or quantifying biomarkers.

In one aspect, the present invention provides a method of identifying and/or quantifying at least one protein, the method comprising: providing at least one protein; labelling at least one amino acid residue of the protein(s) with at least one agent; separating the protein(s) based on a first characteristic; separating the protein(s) based on a second characteristic; and analysing the protein(s) by mass spectrometry, thereby identifying and/or quantifying the at least one protein.

There is also provided a method of identifying and/or quantifying at least one protein from at least two biological materials, the method comprising: providing a first biological material comprising at least one protein and at least one further biological material comprising at least one protein, wherein the protein of the first biological material and the protein of each further biological material are the

same; labelling at least one amino acid residue of the protein(s) with at least one agent; separating the protein(s) based on a first characteristic; separating the protein(s) based on a second characteristic; analysing the protein(s) by mass spectrometry, thereby determining the expression of the at least one protein; and comparing the expression of the protein(s) in the at least two biological materials.

The at least one biological material may be in the form of a sample. The biological materials may be obtained from at least one subject suffering from, or diagnosed with, at least two different disease conditions. The at least two biological materials may be obtained from at least one subject under different treatment conditions. The at least two biological materials may be obtained from the same subject at different time points.

There is also provided a method of screening for at least one drug-responsive protein, the method comprising: administering a drug to at least one subject; obtaining at least one protein from the subject; labelling at least one amino acid residue of the protein(s) with at least one agent; separating the protein(s) based on a first characteristic; separating the protein(s) based on a second characteristic; analysing the protein(s) by mass spectrometry, and determining any variation in the expression of the protein(s), thereby identifying the at least one drug-responsive protein.

There is also provided a method of screening for at least one drug that causes a variation in expression of at least one protein, the method comprising: administering a drug to at least one subject; obtaining at least one protein from the subject; labelling at least one amino acid residue of the protein(s) with at least one agent; separating the protein(s) based on a first characteristic; separating the protein(s) based on a second characteristic; analysing the protein(s) by mass spectrometry, and determining any variation in the expression of the protein(s), thereby identifying the at least one drug.

The agent may be a sulfenyl halide. The agent may be 2-nitrobenzenesulfenyl chloride (NBS). The amino acid residue may be tryptophan. The separating steps may be based on at least one characteristic selected from the group consisting of mass, functional group and isoelectric point. The separating steps may be steps for two-dimensional gel electrophoresis (2DE), The analysing step by mass spectrometry may comprise ionizing the protein(s) and detecting the separated protein(s). The ionizing may be selected from the group consisting of electrospray ionisation (ESI), matrix assisted laser desorption/ionisation (MALDI), surface enhanced laser desorption/ionisation (SELDI), nanospray ionisation, atmospheric pressure chemical ionisation (APCI), chemical ionisation (CI), electron impact (EI), fast atom bombardment (FAB), field desorption / field ionisation (FD/FI) and thermospray ionisation (TSP). The analysing may be by tandem mass spectrometry. The tandem mass spectrometry may use quadrupole-time-of-flight and/or ion trap geometries. The

protein(s) may be biomarkers. The separating steps may be followed by digesting the protein(s) before the analyzing step. The digesting may be by at least one enzyme. The digesting may be performed on a membrane.

There is also provided a kit comprising at least one isotope of a sulfenyl halide compound. The sulfenyl halide compound may be 2-nitrobenzenesulfenyl chloride (NBS). The compound may be for labelling of biomolecules. The kit may be for identifying and/or quantifying at least one protein. The kit may further comprise packaging and information pertaining to the use of the compound for purposes such as diagnosis of disease states.

The present invention utilizes sulfenyl halide agents or compounds. The sulfenyl halide agents or compounds of this invention are generally defined according to the following formula

R_1 --S--X (I)

where S is a sulfur atom, X is a halogen atom, and R_1 is a carbon-based moiety. Preferred halogen atoms include chlorine, bromine, and fluorine, with chlorine being the most preferred halogen. The carbon-based moiety can include any monovalent structure known in the field of organic chemistry. Preferably, the monovalent structure is neutral or non-reactive toward a biomolecule chain end. Under the present invention, certain atoms of the sulfenyl halide, such as carbon atoms, may be isotopes to allow differentiation

of the same compound based on at least one difference in the isotope of the compound such as the difference in mass of the isotopes of the compound.

In particular, there is provided a novel strategy using a sulfenyl halide agent such as in combination with 2DE and MS for quantitative proteome analysis (2DE/MS) to labeling technology. More in particular, the sulfenyl halide agent is 2-nitrobenzenesulfenyl chloride (NBS) (Kuyama et al, 2003). NBS labeling involves the binding of light and heavy NBS isotope moieties to tryptophan residues, resulting in a 6 Da mass difference between light and heavy NBS labeled tryptophan peptides.

Mass spectrometry (MS) separate molecules based on their mass-to-charge (m/z) ratios. In an MS system, molecules of interest are ionized, separated in an analyzer based on their m/z ratios and then detected by a detector. The results may be displayed in the form of an m/z spectrum. The MS system usually includes a data analysis sub-system for the analysis of the data.

Ionisation of the molecules may be achieved by any known method in the art. For example, ionisation of the molecules may be achieved by the following methods, including: Atmospheric Pressure Chemical Ionisation (APCI), Thermospray Ionisation (TSP), Chemical Ionisation (CI), Electron Impact (EI), Electrospray Ionisation (ESI), Fast Atom Bombardment (FAB), Field Desorption

/ Field Ionisation (FD/FI), Surface Enhanced Laser Desorption Ionisation (SELDI) and Matrix Assisted Laser Desorption Ionisation (MALDI).

The analyzer may utilize different principles for the analysis of the ions, including quadrupoles, magnetic sectors, and both Fourier transform, quadrupole ion traps and time-of-flight (TOF) analyzers. Two or more analyzers may be arranged in tandem. Commonly used tandem geometries include quadrupole-quadrupole, magnetic sector-quadrupole, and quadrupole-time-of-flight geometries.

A person skilled in the art will be able to select different ionisation, analysis and/or detection methods to best study a biomolecule of interest such as proteins. When using MS for proteins, relative differences in protein quantity may be determined by measuring the ratio intensity of these light/heavy paired NBS peptide peaks in a MS spectrum. However, compared to ICAT, the NBS reagent possesses several potential advantages. First, the size of the NBS molecule is only 153 Da. This relatively small molecule is less likely to cause interference with peptide ionisation, leading to easier interpretation of mass spectra. Second, the NBS tag does not interfere with the efficiency of collision induced dissociation in MS/MS experiments, and thus no fragments are formed due to the labeling moiety during MS/MS analysis (Kuyama et al., 2003). Finally, as tryptophan is the least abundant residue in proteins with a wide

distribution across protein lengths, the mass spectra obtained using NBS labeling are less analytically complex than ICAT.

In this study, we have applied the NBS/2DE/MS strategy to study proteomic alterations in MCF-7 breast cancer cells induced by tamoxifen (TAM) treatment. TAM is the first line hormonal treatment for 4 estrogen receptor positive (ER+) breast cancer, and although previous reports have described the mRNA gene expression patterns in breast cancers and cell lines associated with TAM treatment (Saji et al., 2005). This is, to the knowledge of the present inventors, the first report investigating TAM-induced proteomic alterations in breast cancer. Using NBS/2DE/MS, 88 protein spots that were quantitatively altered between TAM-treated and control MCF-7 cells were selected, of which 44 could be successfully identified and quantified by NBS labeling. By comparing the protein quantification measurements produced by NBS/2DE/MS with conventional 2DE techniques, we found that NBS/2DE/MS represents a more reliable approach for cellular protein quantification, and further confirmed that the NBS/2DE/MS method is both technically and biologically reproducible.

Of 88 protein spots that significantly changed upon TAM treatment, 44 protein spots were successfully identified with NBS paired peptides, including GRP78, CK18, and CK19. To validate the NBS/2DE/MS results, we used immunoblotting to confirm GRP78, CK19, and PA2G4 as bona-fide TAM regulated proteins, and in two independent breast cancer patient cohorts we

demonstrate that PA2G4 expression can serve as a novel prognostic factor for disease free survival. To the best of inventors' knowledge, this is the first report describing the proteomic changes in breast cancer cells induced by TAM, the most commonly used selective estrogen receptor modulator (SERM). Our results indicate that NBS/2DE/MS may represent a more reliable approach for cellular protein quantitation than conventional 2DE approaches.

While the examples show how the method of the present invention is applied to different treatment conditions, it will be appreciated by one skilled in the art that the method of the present invention may also be used to compare proteomic changes between different disease conditions or to track changes through a course of treatment as well.

Having now generally described the invention, the same will be more readily understood through reference to the following examples which are provided by way of illustration, and are not intended to be limiting of the present invention.

Examples

Materials and Methods

Example 1 – Chemicals and Materials

MCF-7 breast cancer cells were obtained from the American Type Culture Collection (Manassas, VA, USA). Tamoxifen (TAM), chicken ovalbumin, streptomycin, L-glutamine, penicillin, 4-hydroxycinnamic acid (CHCA), 3-

hydroxy-4-nitrobenzoic acid (HNBA), 2,5-dihydroxybenzoic acid (DHB), protease inhibitor cocktail (comprising 4-(2-aminoethyl)benzenesulfonyl fluoride (AEBSF), pepstatinA, E-64, bestatin, leupeptin, and aprotinin), phenyl methyl sulfonyl fluoride (PMSF), sodium deoxychlate and sodium orthovanadate were from Sigma (St Louis, MO, USA). The sulfenyl halide agent 2-nitrobenzenesulfenyl chloride (NBS, heavy and light isotopes) were from Shimadzu Biotech (Kyoto, Japan). Tris (2-carboxyethyl) phosphine (TCEP) hydrochloride was from Pierce (Rockford, IL, USA). Deep Purple fluorescent dye, dithiotritol (DTT) and horseradish peroxidise conjugated secondary antibodies were from GE HealthCare (Uppsala, Sweden). Ethylene diamine tetra acetic acid disodium salt dihydrate (EDTA) and polyvinylidene difluoride membrane (PVDF) were from Bio-Rad (Hercules, CA, USA). C18 ZipTips were from Millipore (Bedford, MA, USA). Sequencing grade modified trypsin was from Promega (Madison, WI, USA). Dulbecco modified Eagle medium 5 (DMEM) was from Gibco (Grand Island, NY, USA). Dextran charcoal-stripped fetal bovine serum (FBS) was from HyClone Laboratories (Pittsburg, PA, USA). Anti-vinculin and anti-PA2G4 antibodies were from Upstate Biotechnology (VA, USA). Anti-cytokeratin 19 and anti-GRP78 antibodies were from Santa Cruz Biotechnology (Santa Cruz Inc, CA, USA). Oligonucleotide primers used in cDNA synthesis and RT-PCR were synthesized from Sigma Proligo (Sigma, Singapore).

Example 2 - Cell Culture and TAM Treatment of MCF-7 Cells

MCF-7 breast cancer cells were cultured in DMEM supplemented with 10% FBS, 100 U/mL penicillin, 100 U/mL streptomycin, and 2 mM Lglutamine. Before TAM treatment, cells were washed in PBS and maintained in phenol red free DMEM with 5 % Dextran charcoal-stripped FBS for 24 hrs. Subsequently, cells were treated with 10 μM TAM at 40-50% confluence and harvested at 24 and 48 hrs. As control, sister cultures were treated with an equivalent volume of the vehicle (0.1% ethanol).

Example 3 - Protein Extraction and NBS Labeling

TAM-treated and control MCF-7 cells were lysed in 2DE lysis buffer (7M urea, 2M thiourea, 4% CHAPS, 1mM PMSF, 50 μg/mL Dnase I, 50 μg/mL Rnase I and protease inhibitor cocktail) and the extracted proteins buffer exchanged into 8M urea and 5mM EDTA using a 2D-sample clean up kit (Bio-Rad). Proteins were quantitated by the Bradford method, Bradford, (1976) and 100 μg of each protein lysate was labeled in the dark with a 20-fold molar excess of 12C (light) or 13C (heavy) 2-nitrobenzenesulfenyl chloride (NBS) in 25 μL of acetic acid with shaking. In NBS labeling, a 12C or 13C NBS moiety is selectively introduced onto tryptophan residues, which imparts a 6 Da mass difference between light and heavy NBS labeled tryptophan peptides. Using mass spectroscopy (MS), the ratio intensity of the paired peptide peaks can then be determined, providing a relative quantitation of the respective light and heavy populations (Figure 1A) (Kuyama et al., 2003).

After labeling, the two protein mixtures were combined and the unlabeled NBS moieties removed by a Sephadex LH-20 column (GE HealthCare). Notably, we have previously shown that the use of a Sephadex LH-20 column to remove excess NBS label does not significantly interfere with protein recovery (see Figure 2 in Matsuo et al.2006), and that protein recovery is robust regardless of whether the samples are denatured with either urea or GdnHCl. The sample was dried with a vacuum concentrator and resuspended in 50 µL of 50mM Tris buffer (pH 8.5) containing 0.1% SDS and reduced with 4mM TCEP for 30 min at 37°C, followed by alkylation with 10mM iodoacetetamide for 45 min at room temperature in the dark. The lysate was then buffer exchanged into 2DE lysis buffer (7M urea, 2M thiourea and 4% CHAPS).

Example 4 - Two-dimensional Gel Electrophoresis. Prior to the first dimensional separation, each IPG strip (18 cm, pH 4-7, GE HealthCare) was rehydrated with 340 µL of rehydration buffer (7M urea, 2M thiourea, 4% CHAPS, 0.5% IPG buffer pH 4-7 and 20mM DTT) for 8 hrs. The first dimensional separation was performed on the IPGphor IEF system (GE HealthCare). Protein lysates of 100 µg in 80 µL of lysis buffer, 20mM DTT and 0.5% IPG buffer, pH 4-7, were applied through cup loading anodically. The strips were then focused at 20°C: 200V for 3 hr; 500V for 1 hr; from 500 to 8000V for 7 hr; and 8000V until a total of 67,000 Vh was reached. After focusing, the IPG strips were equilibrated in two steps: (1) 15 min in 50mM Tris-HCl, pH 8.8, 6M urea, 30% glycerol, 2%

SDS, 1% DTT and a trace of bromophenol blue; (2) 15 min in a similar solution containing 2.5% iodoacetamide instead of DTT.

Once equilibrated, the strips were transferred onto 10% isocratic polyacrylamide gels (18cm x 20cm x 0.75mm). The IPG gels were then sealed with 0.5% w/v agarose in running buffer (25mM Tris-HCl, pH8.3, 192mM glycine and 0.1% SDS). The second dimensional separation was then performed using the Protean II XL system (Bio-Rad). The SDS-PAGE was carried out at 17°C with 8mA/gel for 1 hr followed by 24mA/gel constant current, until the dye front reached the bottom of the gel. The gels were then labeled with the Deep Purple fluorescence dye according to the manufacturer's protocol (GE HealthCare). The gels were digitized using a FX Molecular Imager (Bio-Rad), and image analysis performed using PDQuest 7.3 image analysis software (Bio-Rad).

Example 5 - Protein Identification by Mass Spectrometry

Fluorescence labeled protein gel spots were excised and in-gel trypsin digested using the Xcise robotic system (Shimadzu Biotech, Kyoto, Japan). The automatic operation involved washing the gel pieces with water twice, shrinking with 100% acetonitrile (ACN), and drying prior to digestion. Thirty microliters of trypsin (3.33ng/µL in 50mM ammonium bicarbonate, pH 8.5) was added to each gel plug and the digestion was performed overnight at 30°C with shaking.

The samples were then cleaned and concentrated with C18 ZipTips. Finally, the peptide mixtures were eluted onto the MALDI sample plate with 1 µL of 50% ACN/0.5% TFA and mixed with 1 µL matrix containing 10mg/mL α-cyano-4-hydroxycinnamic acid (CHCA), 2,5-dihydroxybenzoic acid (DHB), and 3-hydroxy-4-nitrobenzoic acid (HNBA) in 50% ACN/0.5% TFA prior to MS analysis. This is a novel matrix formulation based on Matsuo et al., 2006 instead of the regular CHCA/DHB matrix (see Laugesen et al, 2003). MALDI-TOF MS analyses were performed using the AXIMA-CFR plus mass spectrometer (Shimadzu Corporation, Kyoto, Japan and Kratos Analytical, Manchester, UK), under the following settings: nitrogen laser (337 nm); reflectron mode; detection of positive ions. The acceleration potential was set to 35 kV using a gridless-type electrode. MALDI-TOF MS spectra were acquired in manual mode, from m/z 800 to 3000 and internally calibrated with two trypsin autolysis peaks (m/z 842.51 and 2211.10). Peak lists from peptide mass mapping spectra were automatically extracted and submitted to an in-house Mascot server (Matrix Science, London, UK) to search against UniProt and NCBInr databases.

To assist searching for NBS-labeled peptides, two modification values of 12CNBS (+153 Da) and 13CNBS (+159 Da) were added to the Mascot server as variable modifications. The NBS-labeled paired peaks were manually selected from the TOF MS spectrum and the relative quantitation was achieved by calculating the volume ratio of each NBS labeled peak pairs using the

Kompact software (Shimadzu/Kratos). MS/MS peptide sequencing was performed using the AXIMA-QIT MALDI quadrupole ion trap time-of-flight mass spectrometer (Shimadzu Corporation, Kyoto, Japan and Kratos Analytical, Manchester, UK) equipped with a 337 nm nitrogen laser. One microliter of tryptic digested sample was deposited on to the MALDI plate and mixed with 1 µL of DHB (10 mg/mL in 50% ACN/0.5% TFA). The TOF spectrum was externally calibrated using fullerite pre-deposited onto the MALDI sample stage. All spectra were acquired with standard instrument settings for optimum transmission at medium and high masses. Data acquisition and processing were performed using the Kompact software (Shimadzu/Kratos).

Alternatively, instead of excising the proteins from the gel, the proteins may be blotted from the gel onto a membrane, subjected to on-membrane proteolytic digestion using one or more protease applied by a chemical inkjet printer (Shimadzu CHIP 1000), mixed with MS matrix material and placed into an MS system for MALDI-TOF MS analysis (Ohtsu et al, 2005; Nakanishi et al, 2005).

Example 6 - Statistical analysis

To investigate the biological and experimental variations associated with NBS/2DE/MS, and to compare protein quantities between NBS/2DE/MS and conventional 2DE analysis, we applied a series of Pearson's correlation tests to assess the similarity of the two datasets under investigation. For the correlation tests, to determine the significance of any observed correlation value, a

bootstrap permutation test was performed by scrambling the order of the proteins in the two datasets being compared, and re-calculating the correlation coefficient for the randomized datasets. This procedure was repeated 10,000 times, the frequency at which the randomized coefficients exceeded the observed value was computed, and this was reflected as an empirical p-value.

Example 7 - Western Blotting Analysis and Semi-Quantitative RT-PCR Analysis

Whole cell lysates were prepared by lysing TAM-treated and control cells in RIPA buffer (0.15 M NaCl, 50 mM Tris-HCl (pH 7.4), 1 mM EDTA, 1% Triton X-100, 1% sodium deoxycholate, 0.1% SDS) containing 1 mM PMSF and 1 mM sodium orthovanadate. The lysates were quantitated with the Bradford reagent (Bradford, 1976) using BSA as a standard. Equal amounts of the cell lysate (50 µg) were run in 12% SDS-polyacrylamide gel and electroblotted onto a PVDF membrane. The blots were probed with anti-vinculin, anti-cytokeratin 19, anti-GRP78 and anti-PA2G4 antibodies respectively. After incubating the membrane with horseradish peroxidase-conjugated secondary antibody, the reactivity was visualized on X-ray films (Kodak) using a chemiluminescent detection kit (Bio-Rad).

For cDNA synthesis and RT-PCR, total RNA was extracted from TAM treated MCF-7 cells using Trizol reagent. RNA was quantitated and equal quantities of

RNA samples were electrophoresed in a 2% agarose gel to confirm the normalization. Equal quantities of vehicle treated and TAM treated RNA samples were reverse transcribed by Superscript II Reverse transcriptase using oligo-dT primers in the conditions suggested by the manufacturer. RT-PCR was performed from the cDNA using PA2G4 and β-actin specific primers. Sequences of the oligos used in the study were, PA2G4-RTF: CACGTGCCTTCTTCAGTGAG (SEQ ID NO: 56); PA2G4-RTR: ACTCTGGAGGAGGGCCTTTA (SEQ ID NO:57); β-actin-RTF: CGGGAAATCGTGCGTGACATT (SEQ ID NO:58); β-actin-RTR: TGATCTCCTTCTGCATCCTGT (SEQ ID NO:59).

Results

Example 8 - Enhancement of NBS-peptide Signals by a HNBA/CHCA/DHB Matrix.

The major goal of this study was to analyse the feasibility of combining tryptophan-targeted NBS labeling with 2DE and MS to quantify and compare protein expression patterns in two related samples. One challenge of NBS/2DE/MS approach is that NBS labeled peptide ions often exhibit low signal intensity compared to other peptide ions during ionisation, resulting in difficulties for relative quantitation. Recently, we have shown that the addition of 3-hydroxy-4-nitrobenzoic acid (HNBA) to the MALDI matrix can promote NBS

peptide ionisation (Matsuo et al, 2006). This earlier work, however, was limited as it analysed only a single purified protein (lyzozyme). We further tested if HNBA/CHCA/DHB matrices could be used to improve the NBS peptide signals in an MS spectrum for another protein (Human Keratin 18). Compared with the CHCA/DHB matrices alone, the HNBA/CHCA/DHB matrices greatly improved the ionisation of NBS labeled peptides, at similar parameter settings for MS data acquisition (Figure 3). We thus used the HNBA/CHCA/DHB matrices for all subsequent NBS-labeling experiments in this study.

Example 9 - Quantitative Robustness of NBS Labeling

To assess the ability of NBS labeling to accurately quantitate protein concentration differences, we then performed experiments involving purified proteins. Specifically, we created mixtures of purified chicken ovalbumin proteins labeled with either 12C or 13C NBS at different concentration ratios (1:1, 1:2, 1:3, etc), and subjected the NBS-labeled proteins to SDS-PAGE. The ovalbumin bands were excised, digested with trypsin, and the NBS-labeled tryptophan peptides were quantitated. Two NBS peptide pairs (m/z 1734.7/1740.7 and m/z 2012.0/2118.0), whose members were separated by a characteristic 6 Da size difference, were detected in the MALDI-TOF MS spectrum corresponding to the ovalbumin protein (Figure 2). At every concentration ratio, we found across multiple replicate experiments that the relative intensity ratios of both NBS peptide pairs corresponded well to the actual sample concentration ratios, across a 3-fold dynamic range (Figure 2B

and Table 1). Furthermore, there was no significant difference between the first (m/z 1734.7/1740.7) and second (m/z 2012.0/2118.0) peptide pair ratios (t-test, p-values range from 0.1 to 0.7) (Table 1), suggesting that the experimental ratios are independent of peptide identity as long as the peptides are derived from the same protein. There was also no significant difference between the experimental ratios of either NBS peptide to the theoretical expected ratio. This result suggests that the NBS labeling procedure is likely to be quantitatively robust and highly reproducible.

Table 1

Expected ratio (13C/12C)	**Peptide 1 ratio m/z 1734.7/1740.7 (13C/12C) ± SD**	**Peptide 2 ratio m/z 2012.0/2018.0 (13C/12C) ± SD**	**p-value (t-test)**
1	1.22 ± 0.16	1.00 ± 0.06	0.13
2	2.13 ± 0.07	1.83 ± 0.16	0.18
3	2.83 ± 0.38	3.10 ± 0.61	0.66
0.5	0.44 ± 0.01	0.48 ± 0.01	0.11
0.33	0.33 ± 0.08	0.36 ± 0.09	0.71

Table 1. (A) Peptide 1 (LTEWTSSNVMEER, m/z 1734.7/1740.7; SEQ ID NO:52) and Peptide 2 (ELINSWVESQTNGIIR, m/z 2012.0/2018.0; SEQ ID NO:53) were NBS-labeled peptide pairs from chicken ovalbumin used for testing the NBS labeling efficiency. The peak area of each peptide was determined using the Kompact software (Shimadzu/Kratos). Three replicates were made for each ratio test. The p-values were calculated by t-tests comparing the experimental ratios of Peptide 1 and Peptide 2 corresponding to the same expected ratios.

Two NBS peptide pairs (m/z 1734.7/1740.7 and m/z 2012.0/2118.0), whose members were separated by a characteristic 6 Da size difference, were detected in the MALDI-TOF MS spectrum corresponding to the ovalbumin protein (Figure 3A). We found in multiple replicate experiments that the relative intensity ratios of both NBS peptide pairs corresponded well to the actual sample concentration ratios, across a 3-fold dynamic range (Figure 3B and Table 1). The result indicates that the NBS labeling method can be used to reliably quantify proteins with expected errors of 1-11%.

Example 10 – Quantitative NBS/2DE/MS analyses of TAM-treated MCF-7 cells

We then asked if the NBS/2DE/MS approach could be applied to complex biological samples, by profiling MCF-7 breast cancer cells treated with either TAM or vehicle. First, to provide a basis of comparison, the MCF-7 samples

were analysed using a conventional 2DE platform. In this conventional 2DE analysis, triplicate gels for TAM treated and vehicle treated MCF-7 cells were generated, visualized with Deep Purple fluorescence labeling, and PDQuest software used to identify differences in fluorescence label intensities. The resulting 2DE gels showed highly similar protein patterns for samples treated with either TAM or vehicle, with a 85 to 95% match rate across the approximately 340 protein spots in each gel (Figure 12A). Qualitatively, one protein spot was unique in the TAM treated cells, while seven were unique in the vehicle treated cells. At a quantitative level, 27 and 47 protein spots exhibited a >2-fold intensity increase and decrease respectively upon TAM treatment (Figure 12B). Second, the same protein lysates from the TAM-treated and vehicle-treated cells were also subjected to NBS/2DE/MS analysis.

In Table 2 below, the 44 identified NBS protein spots representing 23 different protein species. The 13C/12C ratios were calculated based on the NBS-paired peak intensities. Three proteins were identified with more than one pair of NBS peptides. For them, the 13C/12C ratios were the average of all the paired peptide ratios of the same protein. To test the technical reproducibility, protein lysates from the exact same batch of TAM and vehicle treated MCF-7 cells were re-labelled with NBS, separated on 2DE gels, and re-quantified based on the 13C/12C ratios. Guided by the conventional 2DE gel image analysis, 88 spots were excised from the NBS-labeled 2DE gels, out of which 58 protein spots were positively identified by peptide mass fingerprinting. Spots 1-44 were

proteins detected with NBS-labeled peptide peaks. Among them 44 protein spots were detected with NBS-labeled peptide pairs, which were quantified by calculating 13C/12C peptide volume ratios.

TABLE 2

S/N	Acc No	Protein Name	NBS peptide	Sequence	13C/12C (1)	13C/12C (2)	Mean ± std
1	P05783	K1C18_HUMAN	1135/1141	DWSHYFK (SEQ ID NO:1)	0.15	0.13	0.13 ± 0.01
2	P05783	K1C18_HUMAN	1135/1141	DWSHYFK (SEQ ID NO:2)	0.24	0.56	0.41 ± 0.22
3	P05783	K1C18_HUMAN	1135/1141	DWSHYFK (SEQ ID NO:3)	0.28	0.38	0.32 ± 0.06
4	P05783	K1C18_HUMAN	1135/1141	DWSHYFK (SEQ ID NO:4)	0.54	0.78	0.65 ± 0.17
5	P10809	CH60_HUMAN	1497/1503	TVIIEQSWGSPK (SEQ ID NO:5)	0.45	0.47	0.46 ± 0.01
6	P10809	CH60_HUMAN	1497/1503	TVIIEQSWGSPK (SEQ ID NO:6)	0.95	1.09	1.02 ± 0.09
7	P10809	CH60_HUMAN	1497/1503	TVIIEQSWGSPK (SEQ ID NO:7)	0.36	0.35	0.35 ± 0.01
8	P10809	CH60_HUMAN	1497/1503	TVIIEQSWGSPK (SEQ ID NO:8)	1.09	1.85	1.46 ± 0.53
9	P04792	HSPB1_HUMAN	1094/1100	HTADRWR (SEQ ID NO:9)	0.09	0.08	0.08 ± 0.01
10	P04792	HSPB1_HUMAN	1094/1100	HTADRWR (SEQ ID NO:10)	0.13	0.50	0.32 ± 0.25
11	P04792	HSPB1_HUMAN	1113/1119	DWYPHSR (SEQ ID NO:11)	0.56	0.70	0.63 ± 0.09

12	P30101	PDIA3_HUMAN	935/941	GSNYWR (SEQ ID NO:12)	0.98	1.07	1.03 ± 0.06
13	P30101	PDIA3_HUMAN	935/941	GSNYWR (SEQ ID NO:13)	0.57	0.76	0.66 ± 0.13
14	P30101	PDIA3_HUMAN	935/941	GSNYWR (SEQ ID NO:14)	1.15	1.27	1.21 ± 0.07
15	P11021	GRP78_HUMAN	1583/1589	TWNDPSVQQDIK (SEQ ID NO:15)	1.81	1.16	1.46 ± 0.45
16	P11021	GRP78_HUMAN	1583/1589	TWNDPSVQQDIK (SEQ ID NO:16)	1.21	1.63	1.41 ± 0.29
17	P11021	GRP78_HUMAN	1583/1589	TWNDPSVQQDIK (SEQ ID NO:17)	1.35	0.92	1.13 ± 0.30
18	P11021	GRP78_HUMAN	1583/1589	TWNDPSVQQDIK (SEQ ID NO:18)	1.31	0.84	1.07 ± 0.33
19	P08727	K1C19_HUMAN	1161/1167 1235/1241	IRDWYQK (SEQ ID NO:19) DAEAWFTSR (SEQ ID NO:20)	0.35	0.48	0.41 ± 0.18
20	P08727	K1C19_HUMAN	1161/1167 1235/1242	IRDWYQK (SEQ ID NO:21) / DAEAWFTSR (SEQ ID NO:22)	0.30	0.71	0.51 ± 0.30
21	P08727	K1C19_HUMAN	1161/1167 1235/1242	IRDWYQK (SEQ ID NO:23) / DAEAWFTSR (SEQ ID NO:24)	0.19	0.16	0.17 ± 0.11
22	P08727	K1C19_HUMAN	1161/1167 1235/1242	IRDWYQK (SEQ ID NO:25) / DAEAWFTSR (SEQ ID NO:26)	0.29	0.19	0.24 ± 0.15

23	P08727	K1C19_HUMAN	1161/1167 1235/1242	IRDWYQK (SEQ ID NO:27) / DAEAWFTSR (SEQ ID NO:28)	0.68	0.86	0.77 ± 0.21
24	P30040	ERP29_HUMAN	1090/1096 1218/1224	WAEQYLK (SEQ ID NO:29)/ KWAEQYLK (SEQ ID NO:30)	0.87	1.14	1.01 ± 0.18
25	P30040	ERP29_HUMAN	1223/1229	VGAIQRWLK (SEQ ID NO:31)	NA	0.71	0.71
26	P07237	PDIA1_HUMAN	1355/1361	EADDIVNWLK (SEQ ID NO:32)	0.21	0.29	0.25 ± 0.05
27	P07237	PDIA1_HUMAN	1355/1361	EADDIVNWLK (SEQ ID NO:33)	0.43	0.43	0.62 ± 0.25
28	P08238	HS90B_HUMAN	2409/2415	HNDDEQYAWESSAGGSFTVR (SEQ ID NO:34)	0.98	0.37	0.67 ± 0.42
29	P08238	HS90B_HUMAN	2409/2415	HNDDEQYAWESSAGGSFTVR (SEQ ID NO:35)	0.32	0.47	0.39 ± 0.10
30	P05787	K2C8_HUMAN	1183/1189	WSLLQQQK (SEQ ID NO:36)	0.73	0.93	0.83 ± 0.14
31	P05787	K2C8_HUMAN	1183/1189	WSLLQQQK (SEQ ID NO:37)	0.92	0.87	0.89 ± 0.03
32	P52907	CAZA1_HUMAN	946/952	NFWNGR (SEQ ID NO:38)	0.94	0.88	0.91 ± 0.04
33	P15311	EZRI_HUMAN	1000/1006 1328/1334	GFPTWLK (SEQ ID NO:39) IQVWHAEHR (SEQ ID NO:40)	0.27	0.38	0.32 ± 0.13

34	P48643	TCPE_HUMAN	1891/1897	WVGGPEIELIAIATGGR (SEQ ID NO:41)	0.61	0.94	0.77 ± 0.23
35	Q9BQE3	TBA6_HUMAN	2482/2488	AFVHWYVGEGMEEGEFSEAR (SEQ ID NO:42)	0.37	0.27	0.32 ± 0.07
36	P07437	TBB2_HUMAN	1850/1856	NSSYFVEWIPNNVK (SEQ ID NO:43)	0.32	0.24	0.27 ± 0.05
37	Q9UQ80	PA2G4_HUMAN	2080/2086	LVKPGNQNTQVTEAWNK (SEQ ID NO:44)	0.20	0.19	0.19 ± 0.01
38	P55072	TERA_HUMAN	1482/1488	WALSQSNPSALR (SEQ ID NO:45)	0.90	1.12	1.01 ± 0.15
39	P25786	PSA1_HUMAN	1932/1938	NQYDNDVTVWSPQGR (SEQ ID NO:46)	0.38	0.53	0.45 ± 0.10
40	Q02790	FKBP4_HUMAN	1789/1795	VFVHYTGWLLDGTK (SEQ ID NO:47)	NA	0.21	0.21
41	P55809	SCOT_HUMAN	1518/1524	YGDLANWMIPGK (SEQ ID NO:48)	NA	2.61	2.61
42	P04181	OAT_HUMAN	2253/2259	WLAVDYENVRPDIVLLGK (SEQ ID NO:49)	1.08	0.87	0.97 ± 0.15
43	P78371	TCPB_HUMAN	1444/1450	IHPQTIIAGWR (SEQ ID NO:50)	0.38	0.45	0.41 ± 0.04

44	P30041	PRDX6_HUMAN	2049/2055	LIALSIDSVEDHLAWSK (SEQ ID NO:51)	0.29	0.52	0.41 ± 0.15

Table 2. The 44 identified NBS protein spots representing 23 different protein species. The 13C/12C ratios were calculated based on the NBS-paired peak intensities (Column 13C/12C (1)). Three proteins were identified with more than one pair of NBS peptides. For these proteins, the 13C/12C ratios were the average of all the paired peptide ratios of the same protein. To test the technical reproducibility, protein lysates from the exact same batch of TAM and vehicle treated MCF-7 cells were re-labelled with NBS, separated on 2DE gels, and re-quantified based on the 13C/12C ratios (Column 13C/12C (2)). The mean and SD values are based on combining values from the 13C/12C (1) and 13C/12C (2) columns.

Table 3 (below): Second, the same protein lysates from the TAM-treated and control cells were also subjected to NBS/2DE/MS analysis. Guided by the conventional 2DE gel image analysis, 88 spots were excised from the NBS-labeled 2DE gels. A total of 58 protein spots were positively identified by peptide mass fingerprinting (Figure 4 displays the NBS/2DE reference map with the identified protein spots), of which 44 contained NBS peptide pairs. Some protein spots possessed more than one NBS peptide pair, indicating the presence of multiple tryptophan containing peptides in these proteins. For these proteins, the final NBS ratio was calculated by averaging the ratios of all NBS paired peptides associated with the same protein spot. Using NBS ratio analysis, 21/44 protein spots showed a >2-fold intensity alteration upon TAM treatment, and the remaining 23/44 spots changed less than 2-fold between

TAM and vehicle. In comparison, for the conventional 2DE analysis, 17/44 of these spots were associated with a >2 fold difference, while the remaining 27/44 exhibited a <2 fold difference. The 44 identified NBS labeled protein spots represented 23 different protein species (Table 2), where 10 species, including K1C18_HUMAN (CK18), were represented by multiple isoforms/spots on the 2DE gel.

TABLE 3

Spot No.	Acc. No.	Protein Name	Protein Description	Theoretical *Mw*	Theoretical *pI*	Peptide Coverage %	Peptide Matched	Score	Expectation
1	P05783	K1C18_HUMAN	Cytokeratin-18	47897	5.34	31	16	65	4.14E-03
2	P05783	K1C18_HUMAN	Cytokeratin-18	47897	5.34	32	17	95	4.44E-06
3	P05783	K1C18_HUMAN	Cytokeratin-18	47897	5.34	39	16	95	4.64E-06
4	P05783	K1C18_HUMAN	Cytokeratin-18	47897	5.34	48	21	140	1.31E-10
5	P10809	CH60_HUMAN	60 kDa heat shock protein	61016	5.70	32	15	96	3.36E-06
6	P10809	CH60_HUMAN	60 kDa heat shock protein	61016	5.70	24	14	68	2.03E-03
7	P10809	CH60_HUMAN	60 kDa heat shock protein	61016	5.70	37	15	78	2.07E-04
8	P10809	CH60_HUMAN	60 kDa heat shock protein	61016	5.70	16	8	64	4.75E-03
9	P04792	HSPB1_HUMAN	Heat-shock protein beta-1 (HspB1)	22768	5.98	44	12	79	1.65E-04
10	P04792	HSPB1_HUMAN	Heat-shock protein beta-1 (HspB1)	22768	5.98	44	10	55	4.33E-02
11	P04792	HSPB1_HUMAN	Heat-shock protein beta-1 (HspB1)	22768	5.98	32	9	54	5.46E-02
12	P30101	PDIA3_HUMAN	Protein disulfide-isomerase A3 precursor	56747	5.98	30	14	89	1.54E-05

13	P30101	PDIA3_HUMAN	Protein disulfide-isomerase A3 precursor	56747	5.98	25	11	73	6.12E-04
14	P30101	PDIA3_HUMAN	Protein disulfide-isomerase A3 precursor	56747	5.98	31	14	63	6.12E-03
15	P11021	GRP78_HUMAN	78 kDa glucose-regulated protein precursor	72288	5.07	33	18	102	8.26E-07
16	P11021	GRP78_HUMAN	78 kDa glucose-regulated protein precursor	72288	5.07	36	20	120	1.31E-08
17	P11021	GRP78_HUMAN	78 kDa glucose-regulated protein precursor	72288	5.07	46	26	189	1.65E-15
18	P11021	GRP78_HUMAN	78 kDa glucose-regulated protein precursor	72288	5.07	48	31	231	1.04E-19
19	P08727	K1C19_HUMAN	Cytokeratin-19	44079	5.04	50	24	146	3.29E-11
20	P08727	K1C19_HUMAN	Cytokeratin-19	44079	5.04	47	22	139	1.65E-10
21	P08727	K1C19_HUMAN	Cytokeratin-19	44079	5.04	59	25	185	4.14E-15
22	P08727	K1C19_HUMAN	Cytokeratin-19	44079	5.04	61	28	184	5.21E-15
23	P08727	K1C19_HUMAN	Cytokeratin-19	44079	5.04	69	33	271	1.04E-23
24	P30040	ERP29_HUMAN	Endoplasmic reticulum protein ERp29 precursor	28975	6.77	29	12	63	6.71E-03

25	P30040	ERP29_HUMAN	Endoplasmic reticulum protein ERp29 precursor	28975	6.77	31	11	56	3.00E-02
26	P07237	PDIA1_HUMAN	Protein disulfide-isomerase precursor	57081	4.76	33	19	98	2.03E-06
27	P07237	PDIA1_HUMAN	Protein disulfide-isomerase precursor	57081	4.76	30	15	69	1.57E-03
28	P08238	HS90B_HUMAN	Heat shock protein HSP 90-beta	83081	4.97	20	13	51	9.70E-02
29	P08238	HS90B_HUMAN	Heat shock protein HSP 90-beta	83081	4.97	25	17	62	8.45E-03
30	P05787	K2C8_HUMAN	Cytokeratin-8	53510	5.52	52	23	136	3.29E-10
31	P05787	K2C8_HUMAN	Cytokeratin-8	53510	5.52	30	16	113	6.56E-08
32	P52907	CAZA1_HUMAN	F-actin capping protein alpha-1 subunit	32902	5.45	40	12	62	8.26E-03
33	P15311	EZRI_HUMAN	Ezrin (p81)	69225	5.95	31	26	63	6.71E-03
34	P48643	TCPE_HUMAN	T-complex protein 1	59633	5.45	21	12	56	3.36E-02
35	Q9BQE3	TBA6_HUMAN	Tubulin alpha-6 chain	49863	4.96	42	15	76	3.61E-04
36	P07437	TBB2_HUMAN	Tubulin beta-2 chain	49639	4.78	36	20	118	2.07E-08
37	Q9UQ80	PA2G4_HUMAN	Proliferation-associated protein 2G4	43759	6.13	27	12	56	3.07E-02
38	P55072	TERA_HUMAN	Transitional endoplasmic reticulum ATPase	89135	5.14	33	19	101	1.04E-06

39	P25786	PSA1_HUMAN	Proteasome subunit alpha type 1	29537	6.15	22	7	50	1.44E-01
40	Q02790	FKBP4_HUMAN	FK506-binding protein 4	51641	5.35	30	11	60	1.22E-02
41	P55809	SCOT_HUMAN	Succinyl-CoA:3-ketoacid-coenzyme A transferase 1	56122	7.14	26	7	49	1.77E-01
42	P04181	OAT_HUMAN	Ornithine aminotransferase	48504	6.57	30	13	73	6.12E-04
43	P78371	TCPB_HUMAN	T-complex protein 1	57321	6.02	29	11	90	1.28E-05
44	P30041	PRDX6_HUMAN	Peroxiredoxin 6	24888	6.02	54	14	76	3.00E-04
45	P08559	ODPA_HUMAN	Pyruvate dehydrogenase E1 component alpha subunit	43268	8.35	20	11	67	2.80E-03
46	Q9BX68	HINT2_HUMAN	Histidine triad nucleotide-binding protein 2	17151	9.20	41	4	62	8.85E-03
47	P11142	HSP7C_HUMAN	Heat shock cognate 71 kDa protein	70854	5.37	28	15	62	9.06E-03
48	P06576	ATPB_HUMAN	ATP synthase beta chain	56525	5.26	47	18	120	1.31E-08
49	P08865	RSSA_HUMAN	40S ribosomal protein SA (p40)	32702	4.79	20	5	44	5.09E-01
50	P32119	PRDX2_HUMAN	Peroxiredoxin 2	21747	5.67	27	8	62	7.89E-03
51	Q9NX01	TXN4B_HUMAN	Thioredoxin-like protein 4B	17004	5.63	24	3	33	6.56E+00

52	P30048	PRDX3_HUMAN	Thioredoxin-dependent peroxide reductase	27675	7.67	30	8	56	3.07E-02
53	P05388	RLA0_HUMAN	60S acidic ribosomal protein	34252	5.71	29	5	45	4.24E-01
54	P31040	DHSA_HUMAN	Succinate dehydrogenase [ubiquinone] flavoprotein subunit	72645	7.06	12	5	41	9.93E-01
55	P54284	CACB3_HUMAN	Voltage-dependent L-type calcium channel beta-3 subunit	54498	5.93	17	9	50	1.47E-01
56	P06576	ATPB_HUMAN	ATP synthase beta chain	56525	5.26	35	12	161	1.04E-12
57	P35232	PHB_HUMAN	Prohibitin	29786	5.57	46	10	73	6.26E-04
58	P07741	APT_HUMAN	Adenine phosphoribosyltransferase	19464	5.79	35	5	69	1.65E-03

Example 11 – Comparison of NBS/2DE/MS and 2DE quantitation methods.

The relative protein quantitation by NBS analysis for the 44 NBS-peptide containing proteins was then compared with the ratio intensity of these spots on the conventional 2DE gels. We found the protein quantities delivered by these two approaches to be basically concordant (Figure 5), with a correlation coefficient of 0.51, indicating a statistically significant correlation between these two protein quantitation methods ($p<0.0001$ by bootstrap permutation analysis). However some discrepancies between the 2DE and NBS ratios were also noted.

For example, Spots 20, 21, 22, and 23 (Cytokeratin-19, CK19) exhibited reversed ratios by the two different quantitation methods, with up-regulation predicted by the conventional 2DE gels, and down-regulation by NBS/2DE/MS. To investigate this discrepancy, we hypothesized that while the 2DE intensity ratio might be affected by the co-migration of protein mixtures in the same spot (Froment et al., 2005) the quantitation by NBS labeling might be more reliable as it is based directly on MS peptide peaks which measure only one single protein. It is worth noting that the problem of co-migrating proteins is not an isolated phenomenon, but considered sufficiently widespread to be a key limitation of the 2DE/MS platform for comprehensive and quantitative proteome analysis (Gygi et al., 2000).

Using spot 21 as a representative discrepant spot, we tested this hypothesis in two ways. First, we performed an indirect analysis by checking the 3D image of the 2DE spots using the Bio-Rad PDQuest image analysis software (Figure 6). We found that Spot 21, one of the CK19 isoforms showing discrepancy between NBS/2DE/MS and 2DE quantitation, contained several image peaks in the protein spot, implying co-migration of multiple proteins at this spot position. In contrast, Spot 14 (disulfide isomerase ER-60), one of the spots showing good concordance between NBS/2DE/MS and 2DE quantitation, was associated with a clear single 3D image peak. Second, we analysed the MS spectra associated with protein spot 21 using peptide sequencing.

A peptide mass fingerprinting analysis of the TOF spectrum revealed that apart from peptides belonging to the dominant protein CK19, there were extra unassigned peptides present, likely reflecting contamination from matrices, trypsin, and fluorescence dyes. A subsequent MS/MS analysis of one of the matched peptides at m/z 1222.75 confirmed that it indeed belonged to CK19, but that the unassigned peptide at m/z 1419.85 was instead from Cytokeratin-8 (CK8, Figure 7). Both CK19 and CK8 proteins are predicted to have close Mw and pI values, and both proteins had a number of isoforms displayed on the 2DE gels (Spots 19-23 and 30-31 in Figure 4). Thus, these results show that the differences between the quantitation ratios delivered by NBS/2DE/MS and the conventional 2DE gels are likely to be primarily due to protein co-migration artefacts in the latter, and that NBS/2DE/MS may be a more reliable approach

for protein quantitation studies. We believe that this is a key advantage of the NBS/2DE/MS platform.

One powerful advantage of combining isotope labeling and 2DE is its ability to provide information on post-translationally modified proteins (Smolka et al., 2002). While most of the 23 identified proteins had consistent $^{13}C/^{12}C$ ratios among isoforms, two proteins, heat shock 60 kDa protein (CH60_HUMAN) and disulfide isomerase ER-60 (PDIA_HUMAN), exhibited isoform specific responses to TAM-treatment.

Specifically, we observed four NBS-labeled CH60 isoforms on the 2DE map, all possessing roughly the same molecular masses but different isoelectric points (Spots 5-8 in Figure 4). Based on both NBS labeling quantitation and 2DE image analysis, Spot 7 was observed to be down-regulated after 48 hr TAM treatment, while Spot 8 was up-regulated (Figure 8). These two CH60 isoforms possess shifted isoelectric points, which may reflect the effect of various protein modification events such as phosphorylation (de Graauw et al., 2005; Powell et al., 2002). Indeed, a phosphorylation site can be found in the CH60 primary structure (Rush et al., 2005).

Example 12 – Technical and Biological Reproducibility of NBS/2DE/MS for Complex Protein Lysates. Certain investigators have recently raised concerns regarding the reproducibility of isotope labeling proteomic procedures such as

ICAT. To test the technical reproducibility of the NBS/2DE/MS platform, protein lysates from the exact same batch of TAM and vehicle treated MCF-7 cells were re-labelled with NBS, separated on 2DE gels, and 40 of the previously identified protein spots were digested and re-quantified. For 95% of the proteins, a regression analysis between the original and repeat labelling experiments demonstrated a low experimental variation, with a correlation coefficient (R2) value of 0.71 (Figure 9, R2=0.64 for all proteins). Furthermore, the average coefficient of variation (CV) associated with the technical replicates was 22%, compared with a reported average CV of 23% for 2DE alone and 22.6% for ICAT (isotope-coded affinity tags) technology (Molloy et al, 2005). Taken collectively, this data suggests that the NBS/2DE/MS platform is technically reproducible.

We also investigated the biological reproducibility of the NBS/2DE/MS platform. In this experiment, we treated an independent batch of MCF-7 cells with TAM, subjected the untreated and TAM-treated cells to NBS/2DE/MS, and compared the 13C/12C NBS ratios of 21 protein spots between the original experiment and biological replicate experiment. As expected, the correlation coefficient between the two experiments was lower than the previous technical replicate experiments (R2=0.57 for 95% of proteins), possibly due to the introduction of intrinsic biological noise into the experiment. Taken collectively, these results suggest that NBS/2DE/MS procedure is reasonably reproducible on biological samples.

Example 13 – Validation of Differentially Expressed Proteins by Immunoblotting

Finally, to definitely validate the quantitative information obtained from NBS/2DE/MS, we then conducted immunoblotting experiments for two representative proteins. Total cell lysates from independent batches of MCF-7 cells treated with TAM or vehicle for 24 hr and 48 hrs were separated on 1D gels and immunoblotted with protein specific antibodies (Figure 10). The 78 kDa glucose-regulated protein (GRP78), a multi-drug resistant protein, had four isoforms displayed in the NBS/2DE map (Spots 15-18 in Figure 4).

Based on the NBS (13C /12C) ratios and 2DE quantitation results, the expression of all four isoforms were expected to increase upon 48 hr TAM treatment (Table 2 and Figure 5). Immunoblotting with anti-GRP78 confirmed the up-regulation of this protein even at 24 hr after TAM treatment and this increased expression was stable up to 48 hrs (Figure 10). Similarly, there were five CK19 isoforms identified by 2DE analysis (Spots 19 to 23 in Figure 4). Based on the conventional 2DE image analysis, all but Spot 19 exhibited a significant increase (> 2-fold) upon 48 hr of TAM treatment. However, based on the NBS/2DE/MS quantitation, all the CK19 isoforms showed a clear down-regulation upon TAM treatment (Spots 19 to 23 in Figure 5). Immunoblotting with anti-CK19 antibodies unambiguously validated the down-regulation of the protein upon TAM treatment at 48 hr (Figure 10). This data show that the NBS/2DE-based protein quantitative approach is likely to provide more

accurate quantitative information regarding individual proteins compared to the 2DE only quantitative technology.

Kit

Example 14 – Kit

The present invention also provides a kit comprising at least one isotope of at least one sulfenyl halide compound to allowing labelling of biological samples or biomolecules. For example, the kit may comprise ^{12}C (light) and/or ^{13}C (heavy) 2-nitrobenzenesulfenyl chloride (NBS) for use as a label. The kit may further comprise packaging and information pertaining to the use of the compound for purposes. The labelled biological samples or biomolecule may be used for diagnosis of disease states or treatment conditions from the sample or biomolecule were obtained.

For this invention, inventors applied a tryptophan-targeted NBS/2DE/MS strategy to identify proteomic alterations occurring in MCF-7 cells associated with TAM treatment. TAM is a potent antagonist of estrogen and has well known anti-tumor effects on ER+ breast cancer cells (Treeck et al, 2006). However, despite the fact that MCF-7 is the primary cellular model for all ER+ breast cancers, the proteomic analysis of TAM induced changes in MCF-7 cells has to our knowledge not been previously reported.

The present study enhanced understanding of the molecular effects of TAM on a well known breast cancer cellular model. For example, one of the TAM-regulated proteins was CK19, whose expression has been well documented in breast and other cancers (Gharbi et al., 2002; Uenishi et al., 2003; Ding et al., 2004). Inventors found by NBS/2DE/MS that CK19 was down-regulated upon TAM treatment, and validated this observation by immunoblotting TAM-treated samples with a CK19 specific antibody. Previous reports have shown that that CK19 expression is a predictor of early postoperative recurrence (Uenishi et al., 2003) and in hepatocellular carcinoma (HCC), CK19 protein expression has been associated with pathological progression and is considered as a potential therapeutic target for treating HCC patients with metastases (Ding et al., 2004).

Another TAM-regulated protein, CK18 neoepitope M30, has been shown to be a useful serum marker of apoptotic epithelium in septic patients (Roth et al., 2004). During apoptosis, DNA fragmentation, nuclear condensation and CK18 cleavage are known events, and a neoepitope in CK18 called M30 neoantigen, becomes available at an early caspase cleavage event. Notably, even though TAM is the primary hormonal therapy for breast cancer, a major adverse side effect of TAM therapy is undesirable endometrial proliferation.

Interestingly, a previous report on TAM mediated changes in Ishikawa endometrial adenocarcinoma cell lines using 2DE showed a significant increase in CK18 expression (Shah et al., 2004) which is opposite to our observation in

breast cancer cells. The current observation on selective estrogen receptor modulator (SERM) mediated down-regulation of CK18 in the MCF-7 cell line and the exactly opposite observation in endometrial cancer cell line raises the possibility that CK18 might represent an important tissue-specific target of SERMs. Finally, GRP78 is another protein identified by NBS quantitation as being TAM-regulated.

GRP78 is a known drug resistance effector and is also a stress response chaperon. Lung cancer and neuroblastoma patients with a positive GRP78 expression tended to show a better prognosis and improved survival (Uramoto et al., 2005; Hsu et al., 2005). The observed up-regulation of GRP78 in the TAM treated MCF-7 cells may represent an immediate stress response to the drug treatment, and also could couple cellular stress to the cell death program (Rao et al., 2002; Ding et al., 2004). This is supported by accompanied up-regulation of another ER stress gene, HERP, at the mRNA level upon TAM treatment.

To demonstrate the relevance of our study to breast cancer and clinical outcome, we focused on PA2G4, originally identified as an ERBB3 receptor interacting protein. We performed immunoblotting experiments to confirm the down-regulation of PA2G4 protein in MCF7 cells by TAM treatment.

As shown in Figure 10B, exposure of MCF7 cells to TAM induced a modest but noticeable down-regulation of PA2G4 protein after 24hr, which persisted until 48hr. In an independent experiment, we have also found that the expression of PA2G4 protein is >2 fold higher in MCF-7 (ER+) than in HCC-38 (ER-) and CCD-1059sk (ER-) breast cell lines. We hypothesized that the down-regulation of PA2G4 protein by TAM might be transcriptionally mediated, and proceeded to test this hypothesis in a series of semi-quantitative RT-PCR experiments. Similar to PA2G4 protein, we found that PA2G4 mRNA was clearly down-regulated by TAM exposure at both the 24h and 48h time points (Figure 10B).

Given this transcriptional regulation, we then performed data-mining on two independent microarray data sets of ER+ primary tumors. The first set (GIS; Miller et al, 2005) consisted of data from 201 ER+ breast cancer patients, where 67 patients received only adjuvant hormonal therapy after surgery. The second set (Oxford; Sotiriou et al, 2006) consists of data from 65 ER+ patients, where 33 patients received TAM only and the remaining patients did not receive treatment. By univariate statistical analysis, PA2G4 transcript expression demonstrated a highly significant association with disease-free survival in both data sets (p-values are 1.75x10-5 (GIS) and 0.001 (Oxford) respectively), with patients with high PA2G4 expressing tumors exhibiting poorer clinical outcome. The significant association between PA2G4 expression and survival was also observed in both data sets when the univariate analysis was restricted to only

those patients receiving adjuvant hormonal therapy (p values are 0.002 (GIS) and 0.008 (Oxford) respectively) (Table 4)

A)	Univariate Analysis					B)	Multivariate Analysis			
GIS (all ER+, n = 201)										
	P-value	Regression Coefficient	95% Confidence Interval for regression coefficient				P-value	Regression Coefficient	95% Confidence Interval for regression coefficient	
			Lower	Upper					Lower	Upper
PA2G4	**2E-05**	1.001	1.001	1.002		*PA2G4*	**0.001**	1.001	1.000	1.002
						size	0.015	1.033	1.006	1.061
						p53	0.916	0.960	0.452	2.040
						node	**0.004**			
						grade	0.469			
						age	0.531	1.007	0.986	1.028
GIS (Tamoxifen-treated, n = 67)										
	P-value	Regression Coefficient	95% Confidence Interval for regression coefficient				P-value	Regression Coeffi-cient	95% Confidence Interval for regression coefficient	
			Lower	Upper					Lower	Upper
PA2G4	**0.002**	1.001	1.001	1.002		*PA2G4*	**0.005**	1.001	1.000	1.002
						size	0.796	1.006	0.961	1.053
						p53	0.762	1.202	0.366	3.940
						node	0.294			
						grade	0.320			

						age	**0.039**	1.043	1.002	1.086
Oxford (all ER+, n = 65)										
	P-value	Regression Coefficient	95% Confidence Interval for regression coefficient				P-value	Regress-ion Coefficient	95% Confidence Interval for regression coefficient	
			Lower	Upper					Lower	Upper
PA2G4	**0.001**	1.002	1.001	1.004		*PA2G4*	**0.043**	1.002	1.000	1.004
						grade	0.897	1.048	0.519	2.113
						age	0.974	0.999	0.933	1.069
						size	0.302	1.263	0.810	1.969
Oxford (Tamoxifen-treated, n = 33)										
	P-value	Regression Coefficient	95% Confidence Interval for regression coefficient				P-value	Regression Coefficient	95% Confidence Interval for regression coefficient	
			Lower	Upper					Lower	Upper
PA2G4	**0.008**	1.003	1.001	1.005		*PA2G4*	0.185	1.002	0.999	1.005
						grade	0.709	1.225	0.421	3.563
						age	0.260	0.938	0.838	1.049
						node	0.790	1.418	0.108	18.537
						size	0.447	1.260	0.695	2.282

Table 4. Univariate and multivariate analysis of PA2G4 as a survival associated gene in ER+ breast cancer. The significant association ($p<0.05$) is highlighted in bold.

We then tested if PA2G4 expression might behave as an independent prognostic factor in comparison to other clinical variables, such as tumor grade, patient age, and tumor size. In this multivariate analysis, PA2G4 expression still retained its prognostic significance in both patient cohorts (p values are 0.001 (GIS) and 0.04 (Oxford) respectively) and also in the GIS cohort of patients receiving adjuvant hormonal therapy (p = 0.005) (Table 3B). Thus, in two independent microarray data sets of ER+ breast cancers, high PA2G4 expression in primary tumors is associated with decreased disease-free survival compared to patients with low PA2G4 expressing tumors. These results show that PA2G4 is a novel survival-associated breast cancer gene.

The present inventors have described a novel approach for protein quantitation and/or identification, which, in a particular aspect, combines NBS tryptophan labeling with 2DE/MS. There are several technical issues regarding the NBS/2DE/MS platform worth noting. First, to enhance the NBS-peptide signals within the MS spectrum, we used a novel matrix formulation (HNBA/CHCA/DHB) modified from that previously described (Matsuo et al, 2006). The earlier work, however, analysed only a single purified protein (lyzozyme) and did not address the general applicability of the HNBA matrix towards different protein entities in the context of a complex cellular lysate. According to a particular embodiment of the invention, the NBS-labeling, including the purified ovalbumin and TAM studies, were performed using the HNBA matrix. Using this method, the inventors identified NBS-peptide signals

for over 20 distinct proteins which confirm the utility and flexibility of the HNBA/CHCA/DHB matrix.

Second, in a series of in-vitro experiments, we found that NBS ratios associated with different peptides of the same protein (chicken ovalbumin) were statistically similar, across a wide range of differential concentrations (Table 1). Thus, the experimental ratios associated with the NBS/2DE/MS platform do not appear to be dependent on peptide identity, supporting the reliability of the results.

Third, the NBS/2DE/MS platform is able to provide accurate quantitation of protein spots containing co-migrating proteins. Notably, the problem of co-migrating proteins is considered to be a key limitation of the 2DE/MS platform for comprehensive and quantitative proteome analysis. For example, Gygi et alia showed that despite using very narrow pH range isoelectric focusing 2DE gels (pH 4.9 –5.5), a single silver-stained spot could still harbour the protein products of six different genes, resulting in substantial difficulties for quantitative analysis (Gygi et al, 2000).

The ability of the NBS/2DE/MS platform to handle such areas of protein overlap does show its improved ability to impart useful and accurate information regarding protein regulation. Interestingly, it appears that the NBS/2DE/MS platform have higher sensitivity compared to conventional 2DE/MS. Specifically, of 44 protein spots that were quantitated both by NBS and conventional 2DE,

23/44 spots showed >2 fold regulation by NBS/2DE/MS, while only 17/44 spots showed >2 fold regulation by conventional 2DE. For example, as shown in Figure 5, Spot 36 had a NBS ratio of 3.6, but a 2DE ratio of only 1.07. As such, the protein corresponding to Spot 36 appears to have been mistakenly annotated by conventional 2DE as a TAM-nonresponsive protein, when in reality it was up-regulated by TAM treatment. The specific reasons behind the potential greater sensitivity of the NBS/2DE/MS platform remain to be determined.

Fourth, our results suggest that that NBS/2DE/MS is likely to show reasonable technical reproducibility. Specifically, the average CV associated with the NBS/2DE/MS platform was 22%, which is comparable to the reported average CV of 23% for 2DE alone23 and 22.6% for ICAT (isotope-coded affinity tags) technology (Molloy et al, 2005).

Finally, we note the NBS labeled tryptophan residue can not only serve as the means for protein quantitation, but also as an extra parameter for protein identification. With this parameter, the confidence of peptide mass fingerprinting analysis is greatly enhanced. To demonstrate the ability of the NBS/2DE/MS strategy to identify drug-induced proteomic alterations, we applied this platform to compare MCF-7 cells that had either been treated or untreated with TAM. TAM is a potent antagonist of estrogen and has well known antitumor effects on ER+ breast cancer cells.31 However, despite the

fact that MCF-7 is the primary cellular model for all ER+ breast cancers, the proteomic analysis of TAM induced changes in MCF-7 cells has to our knowledge not been previously reported.

The method according to the present invention is also be useful in understanding the molecular effects of TAM on a well known breast cancer cellular model. Indeed, of the 23 protein species identified as being TAM-regulated by the NBS/2DE/MS platform, only 7 have been previously identified as having an association with either TAM treatment or the estrogen receptor (ER). One of the TAM-regulated proteins was CK19, whose expression has been well documented in breast and other cancers (Gharbi et al, 2002; Uenishi et al, 2003; Ding et al, 2004).

We found by NBS/2DE/MS that CK19 was down-regulated upon TAM treatment, and validated this observation by immunoblotting TAM-treated samples with a CK19 specific antibody. Previous reports have shown that CK19 expression is a predictor of early postoperative recurrence (Uenishi et al, 2003) and in hepatocellular carcinoma (HCC) CK19 protein expression has been associated with pathological progression and is considered as a potential therapeutic target for treating HCC patients with metastases (Ding et al, 2004). In breast cancer, elevated CK19 gene expression in bone marrow cells has also been correlated with clinical outcome (Shammas et al, 2001), which is

consistent with our observation that TAM also down-regulates CK19 mRNA level in MCF-7 cells.

One biological finding in our study was the identification of PA2G4 as a potential novel survival associated gene in ER+ breast cancer. Using NBS/2DE/MS, we identified PA2G4 as a protein that was down-regulated in MCF-7 cells by TAM, and subsequently confirmed that this down-regulation is likely due to transcriptional control of the PA2G4 gene. By analysing two different public domain microarray data sets of ER+ tumors, we found that PA2G4 expression in primary ER+ tumors was significantly associated with disease free survival, in both a univariate and multivariate analysis.

Specifically, patients with high PA2G4 expressing tumors appeared to exhibit significantly poorer clinical outcome compared to patients with low-PA2G4 expressing tumors. The association between PA2G4 expression and clinical outcome was also observed in patients that had either been treated with adjuvant hormonal therapy or without any systematic treatment. From our finding in this study that PA2G4 protein is down-regulated by TAM, these results show that the therapeutic benefits of TAM may be mediated, at least in some part, through the down-regulation of PA2G4 protein. PA2G4 has also been shown to participate in the differentiation of human ErbB receptor-positive breast and prostate cancer cells, and is capable of inducing antiproliferative

effects by interacting with the retinoblastoma (Rb) tumor suppressor protein (Yoo et al, 2000; Pompeia et al, 2004).

In conclusion, according to a particular embodiment of the present invention, most of the proteins identified by NBS based quantitation in the TAM treated breast cancer cells have both known and novel biological and clinical associations. Our results have clearly demonstrated that NBS is a reliable isotope labelling reagent for quantitative proteome analysis, and NBS/2DE/MS is a simple but effective platform to accurately quantify proteins and their isoforms even when protein co-migration occurs.

The identification of drug-responsive biomarkers in complex protein mixtures is an important goal of quantitative proteomics. In general, the method of the present invention labels proteins with an agent and separates them based on two characteristics before analysis with mass spectrometry. According to a particular embodiment, the present invention combines 2-nitrobenzenesulfenyl chloride (NBS) tryptophan labelling with two-dimensional gel electrophoresis (2DE)/mass spectrometry (MS). Lysates from drug-treated and control samples are labelled with light or heavy NBS residues, separated on a common 2DE gel, and protein alterations identified by MS through the differential intensity of paired NBS peptide peaks. Using NBS/2DE/MS, the proteomic alterations induced by tamoxifen (TAM) in MCF-7 breast cancer cells was profiled. Of 88 protein spots that significantly changed upon TAM treatment, 44 protein spots

were successfully identified with NBS paired peptides, including GRP78, CK18, CK19 and PA2G4.

The NBS/2DE/MS alterations exhibited good concordance with conventional 2DE differential protein quantitation, with discrepancies due largely to the co-migration of distinct proteins in the regular 2DE gels. According to a particular embodiment of the present invention, the NBS labelling was both technically and biologically reproducible, and independently validated three proteins, GRP78, CK19 and PA2G4, as bona-fide up and down-regulated proteins during TAM treatment in MCF-7 cells by Western blotting, and further show that in primary breast cancers PA2G4 expression behaves as a novel prognostic factor in two independent patient cohorts. The results indicate that NBS/2DE/MS is a reliable approach for cellular protein quantitation than conventional 2DE approaches.

It will be apparent to a person skilled in the art that the present invention may also be used in veterinary medicine for animals. While specific examples to practice the invention have been provided, it will be appreciated that various modifications and improvements may be made by a person skilled in the art without departing from the spirit and scope of the present invention.

References

Aebersold, R.; Mann, M. (2003) Nature, 422, 198-207.

Bradford, M. M. (1976) Anal. Biochem. 72, 248

Campostrini, N.; Areces, L. B.; Rappsilber, J.; Pietrogrande, M. C.; Dondi, F.; Pastorino, F.; Ponzoni, M.; Righetti, P. G. (2005) Proteomics, 5, 2385-2395.

de Graauw, M.; Tijdens, I.; Cramer, R.; Corless, S.; Timms, J. F.; van de, W. B. (2005) J.Biol.Chem., 280, 29885-29898.

Ding, S. J.; Li, Y.; Tan, Y. X.; Jiang, M. R.; Tian, B.; Liu, Y. K.; Shao, X. X.; Ye, S. L.; Wu, J. R.; Zeng, R.; Wang, H. Y.; Tang, Z. Y.; Xia, Q. C. (2004) Mol Cell Proteomics, 3, 73-81.

Froment, C.; Uttenweiler-Joseph, S.; Bousquet-Dubouch, M. P.; Matondo, M.; Borges, J. P.; Esmenjaud, C.; Lacroix, C.; Monsarrat, B.; Burlet-Schiltz, O. (2005) Proteomics., 5, 2351-2363.

Gharbi, S.; Gaffney, P.; Yang, A.; Zvelebil, M. J.; Cramer, R.; Waterfield, M. D.; Timms, J. F. (2002) Mol.Cell Proteomics., 1, 91-98.

Gygi, S. P.; Corthals, G. L.; Zhang, Y.; Rochon, Y.; Aebersold, R. Proc.Natl.Acad.Sci.U.S.A 2000, 97, 9390-9395.

Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. (1999) Nat.Biotechnol., 17, 994-999.

Haynes, P. A.; Yates, J. R., III (2000) Yeast, 17, 81-87.

Herbert, B. R.; Harry, J. L.; Packer, N. H.; Gooley, A. A.; Pedersen, S. K.; Williams, K. L. (2001) Trends Biotechnol., 19, S3-S9.

Hsu, W. M.; Hsieh, F. J.; Jeng, Y. M.; Kuo, M. L.; Tsao, P. N.; Lee, H.; Lin, M. T.; Lai, H. S.; Chen, C. N.; Lai, D. M.; Chen, W. (2005) J. Int.J.Cancer, 113, 920-927.

Kuyama, H.; Watanabe, M.; Toda, C.; Ando, E.; Tanaka, K.; Nishimura, O. (2003) Rapid Commun.Mass Spectrom., 17, 1642-1650.

Laugesen, S.; Roepstorff, P. (2003) J.Am.Soc.Mass Spectrom., 14, 992-1002.

Matsuo, E.; Toda, C.; Watanabe, M.; Ojima, N.; Izumi, S.; Tanaka, K.; Tsunasawa, S.; Nishimura, O. (2006) Proteomics, 6, 2042-2049.

Miller, L. D.; Smeds, J.; George, J.; Vega, V. B.; Vergara, L.; Ploner, A.; Pawitan, Y.; Hall, P.; Klaar, S.; Liu, E. T.; Bergh, J. Proc.Natl.Acad.Sci.U.S.A 2005, 102, 13550-13555.

Molloy, M. P.; Donohoe, S.; Brzezinski, E. E.; Kilby, G. W.; Stevenson, T. I.; Baker, J. D.; Goodlett, D. R.; Gage, D. A. (2005) Proteomics., 5, 1204-1208.

Nakanishi T, Ohtsu I, Furuta M, Ando E, and Nishimura O. J Proteome Res (2005) 4:743-747

Ohtsu I, Nakanisi T, Furuta M, Ando E, and Nishimura O. J Proteome Res (2005) 4:1391-1396.

Pompeia C, Hodge DR, Plass C, Wu YZ, Marquez VE, Kelley JA, Farrar WL. Cancer Res 64: 3465-3473 (2004)

Powell, D. W.; Rane, M. J.; Chen, Q.; Singh, S.; McLeish, K. R. (2002) J.Biol.Chem., 277, 21639-2821642.

Rao, R. V.; Castro-Obregon, S.; Frankowski, H.; Schuler, M.; Stoka, V.; del Rio, G.; Bredesen, D. E.; Ellerby, H. M. (2002) J.Biol.Chem., 277, 21836-21842.

Roth, G. A.; Krenn, C.; Brunner, M.; Moser, B.; Ploder, M.; Spittler, A.; Pelinka, L.; Sautner, T.; Wolner, E.; Boltz-Nitulescu, G.; Ankersmit, H. J. (2004) Shock, 22, 218-220.

Rush, J.; Moritz, A.; Lee, K. A.; Guo, A.; Goss, V. L.; Spek, E. J.; Zhang, H.; Zha, X. M.; Polakiewicz, R. D.; Comb, M. (2005) J. Nat.Biotechnol., 23, 94-101.

Saji, S.; Kawakami, M.; Hayashi, S.; Yoshida, N.; Hirose, M.; Horiguchi, S.; Itoh, A.; Funata, N.; Schreiber, S. L.; Yoshida, M.; Toi, M. (2005) Oncogene, 24, 4531-4539.

Shah, Y. M.; Basrur, V.; Rowan, B. G. (2004) Mol Cell Endocrinol., 219, 127-139.

Shammas, F. V.; Deak, E.; Nysted, A.; van Eekelen, J. A.; Osland, A.; Heikkila, R. (2001) Anticancer Res, 21, 2099-2106.

Smolka, M.; Zhou, H.; Aebersold, R. (2002) Mol.Cell Proteomics., 1, 19-29.

Sotiriou, C.; Wirapati, P.; Loi, S.; Harris, A.; Fox, S.; Smeds, J.; Nordgren, H.; Farmer, P.; Praz, V.; Haibe-Kains, B.; Desmedt, C.; Larsimont, D.; Cardoso, F.; Peterse, H.; Nuyten, D.; Buyse, M.; Van de Vijver, M. J.; Bergh, J.; Piccart, M.; Delorenzi, M. J Natl.Cancer Inst. 2006, 98, 262-272.

Treeck, O.; Wackwitz, B.; Haus, U.; Ortmann, O. (2006.) Gynecol.Oncol. –in press.

Uenishi, T.; Kubo, S.; Yamamoto, T.; Shuto, T.; Ogawa, M.; Tanaka, H.; Tanaka, S.; Kaneda, K.; Hirohashi, K. (2003) Cancer Sci., 94, 851-857.

Uramoto, H.; Sugio, K.; Oyama, T.; Nakata, S.; Ono, K.; Yoshimastu, T.; Morita, M.; Yasumoto, K. (2005) Lung Cancer, 49, 55-62.

Vuong, G. L.; Weiss, S. M.; Kammer, W.; Priemer, M.; Vingron, M.; Nordheim, A.; Cahill, M. A. (2000) Electrophoresis, 21, 2594-2605.

Yoo, J. Y.; Wang, X. W.; Rishi, A. K.; Lessor, T.; Xia, X. M.; Gustafson, T. A.; Hamburger, A. W. (2000) Br.J.Cancer, 82, 683-690.

Claims

1. A method of identifying and/or quantifying at least one protein, the method comprising:

(a) providing at least one protein;

(b) labelling at least one amino acid residue of the protein(s) with at least one agent;

(c) separating the protein(s) based on a first characteristic;

(d) separating the protein(s) based on a second characteristic; and

(e) analysing the protein(s) by mass spectrometry, thereby identifying and/or quantifying the at least one protein.

2. A method of identifying and/or quantifying at least one protein from at least two biological materials, the method comprising:

(a) providing a first biological material comprising at least one protein and at least one further biological material comprising at least one protein, wherein the protein of the first biological material and the protein of each further biological material are the same;

(b) labelling at least one amino acid residue of the protein(s) with at least one agent;

(c) separating the protein(s) based on a first characteristic;

(d) separating the protein(s) based on a second characteristic;

(e) analysing the protein(s) by mass spectrometry, thereby determining the expression of the at least one protein; and

(f) comparing the expression of the protein(s) in the at least two biological materials.

3. The method according to claim 2, wherein at least one biological material is in the form of a sample.

4. The method according to any one of claims 1 to 3, wherein the agent is a sulfenyl halide.

5. The method according to any one of claims 1 to 4, wherein the agent is 2-nitrobenzenesulfenyl chloride (NBS).

6. The method according to any one of claims 1 to 5, wherein the amino acid residue is tryptophan.

7. The method according to any one of claims 2 to 6, wherein the biological materials are obtained from at least one subject suffering from, or diagnosed with, at least two different disease conditions.

8. The method according to any one of claims 2 to 6, wherein the at least two biological materials are obtained from at least one subject under different treatment conditions.

9. The method according to any one of claims 2 to 6, wherein the at least two biological materials are obtained from the same subject at different time points.

10. The method according to any one of claims 1 to 9, wherein the separating steps (c) and (d) are based on at least one characteristic selected from the group consisting of mass, functional group, charge and isoelectric point.

11. The method according to any one of claims 1 to 10, wherein the separating steps (c) and (d) are steps for two-dimensional gel electrophoresis (2DE).

12. The method according to any one of claims 1 to 11, wherein step (e) analysing by mass spectrometry comprises ionizing the protein(s) and detecting the separated protein(s).

13. The method according to claim 12, wherein the ionizing is selected from the group consisting of electrospray ionisation (ESI), matrix assisted

laser desorption/ionisation (MALDI), surface enhanced laser desorption/ionisation (SELDI), nanospray ionisation, atmospheric pressure chemical ionisation (APCI), chemical ionisation (CI), electron impact (EI), fast atom bombardment (FAB), field desorption / field ionisation (FD/FI) and thermospray ionisation (TSP).

14. The method according to any one of claims 1 to 13, wherein step (e) analysing is by tandem mass spectrometry.

15. The method according to claim 14, wherein the tandem mass spectrometry uses quadrupole-time-of-flight and/or ion-trap geometries.

16. A method of screening for at least one drug-responsive protein, the method comprising:

(a) administering a drug to at least one subject;

(b) obtaining at least one protein from the subject;

(c) labelling at least one amino acid residue of the protein(s) with at least one agent;

(d) separating the protein(s) based on a first characteristic;

(e) separating the protein(s) based on a second characteristic;

(f) analysing the protein(s) by mass spectrometry, and

(g) determining any variation in the expression of the protein(s), thereby identifying the at least one drug-responsive protein.

17. A method of screening for at least one drug that causes a variation in expression of at least one protein, the method comprising:

(a) administering a drug to at least one subject;

(b) obtaining at least one protein from the subject;

(c) labelling at least one amino acid residue of the protein(s) with at least one agent;

(d) separating the protein(s) based on a first characteristic;

(e) separating the protein(s) based on a second characteristic;

(f) analysing the protein(s) by mass spectrometry, and

(g) determining any variation in the expression of the protein(s), thereby identifying the at least one drug.

18. The method according to any one of claims 16 to 17, wherein the agent is a sulfenyl halide.

19. The method according to any one of claims 16 to 18, wherein the agent is 2-nitrobenzenesulfenyl chloride (NBS).

20. The method according to any one of claims 16 to 19, wherein the amino acid residue is tryptophan.

21. The method according to any one of claims 16 to 20, wherein the separating steps (d) and (e) are based on at least one characteristic

selected from the group consisting of mass, functional group and isoelectric point.

22. The method according to any one of claims 16 to 21, wherein the separating steps (d) and (e) are steps for two-dimensional gel electrophoresis (2DE),

23. The method according to any one of claims 16 to 22, wherein step (f) analysing by mass spectrometry comprises ionizing the protein(s) and detecting the separated protein(s).

24. The method according to any one of claim 23, wherein the ionizing is selected from the group consisting of electrospray ionisation (ESI), matrix assisted laser desorption/ionisation (MALDI), surface enhanced laser desorption/ionisation (SELDI), nanospray ionisation, atmospheric pressure chemical ionisation (APCI), chemical ionisation (CI), electron impact (EI), fast atom bombardment (FAB), field desorption / field ionisation (FD/FI) and thermospray ionisation (TSP).

25. The method according to any one of claims 16 to 24, wherein the) analysing is by tandem mass spectrometry.

26. The method according to claim 25, wherein the tandem mass spectrometry uses quadrupole-time-of-flight and/or ion trap geometries.

27. The method according to any one of claims 1 to 26, wherein the protein(s) is a biomarker.

28. The method according to any one of claims 1 to 27, wherein the separating steps are followed by digesting the protein(s) before the analyzing step.

29. The method according to claim 28, wherein the digesting is by at least one enzyme.

30. The method according to any one of claims 28 to 29, wherein the digesting is performed on a membrane.

31. A kit comprising at least one isotope of a sulfenyl halide compound.

32. The kit according to claim 31, wherein the sulfenyl halide compound is 2-nitrobenzenesulfenyl chloride (NBS).

33. The kit according to claim 31 or 32, wherein the compound is for labelling of biomolecules.

34. The kit according to any one of claims 31 to 33, wherein the kit is for identifying and/or quantifying at least one protein.